

(12) **United States Patent**
Marvit

(10) **Patent No.:** **US 9,247,205 B2**
(45) **Date of Patent:** **Jan. 26, 2016**

(54) **SYSTEM AND METHOD FOR EDITING
RECORDED VIDEOCONFERENCE DATA**

2003/0169330 A1* 9/2003 Ben-Shachar et al. 348/14.09
(Continued)

(75) Inventor: **David L. Marvit**, San Francisco, CA
(US)

OTHER PUBLICATIONS

(73) Assignee: **Fujitsu Limited**, Kawasaki-shi (JP)

Gail Jensen, "Video Conferencing with Archival Notes," U.S. Appl. No. 12/012,044, 26 pages, Filed Jan. 30, 2008.
Thomas H. Hess et al., "Systems and Methods for Conferencing Among Governed and External Participants," U.S. Appl. No. 11/005,545, 85 pages, Filed Dec. 6, 2004.
(Continued)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1138 days.

(21) Appl. No.: **12/872,319**

Primary Examiner — Hung Dang

(22) Filed: **Aug. 31, 2010**

Assistant Examiner — Girumsew Wendmagegn

(74) *Attorney, Agent, or Firm* — Baker Botts L.L.P.

(65) **Prior Publication Data**

US 2012/0051719 A1 Mar. 1, 2012

(51) **Int. Cl.**

H04N 9/80 (2006.01)

H04N 7/15 (2006.01)

G11B 27/034 (2006.01)

G10L 15/26 (2006.01)

G10L 17/00 (2013.01)

G10L 25/48 (2013.01)

(52) **U.S. Cl.**

CPC **H04N 7/155** (2013.01); **G11B 27/034**
(2013.01); **G10L 15/26** (2013.01); **G10L 17/00**
(2013.01); **G10L 17/005** (2013.01); **G10L 25/48**
(2013.01)

(58) **Field of Classification Search**

CPC H04N 7/15; H04N 7/155; H04N 21/4788
USPC 386/235, 285, 278; 348/14.08
See application file for complete search history.

(56) **References Cited**

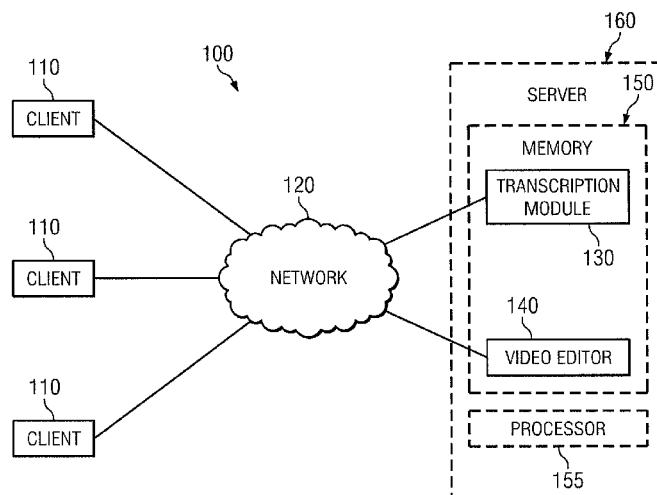
U.S. PATENT DOCUMENTS

7,046,779 B2 5/2006 Hesse
7,185,054 B1 2/2007 Ludwig et al.

(57) **ABSTRACT**

In a method embodiment, a method for automatically editing data recorded during a videoconference includes accessing a plurality of video data streams. Each video data stream records a view of at least one of a plurality of human participants of the videoconference. The view recorded by each video data stream is different from the view recorded by each other video data stream. The method further includes determining, using one or more processors executing logic, that one of the plurality of video data streams recorded a view of a first one of the plurality of participants while the first one of the plurality of participants made a first statement. In addition the method includes determining, using one or more processors executing logic, that one of the plurality of video data streams recorded a view of a second one of the plurality of participants while the second one of the plurality of participants made a second statement after the first one of the plurality of participants made the first statement. An edited video data stream is generated using the plurality of video data streams. The edited video data stream comprises a transition that switches from the view of the first one of the plurality of participants to the view of the second one of the plurality of participants. The transition is timed such that when the edited video data stream is played the transition occurs before the commencement of the second statement.

14 Claims, 2 Drawing Sheets



(56)

References Cited

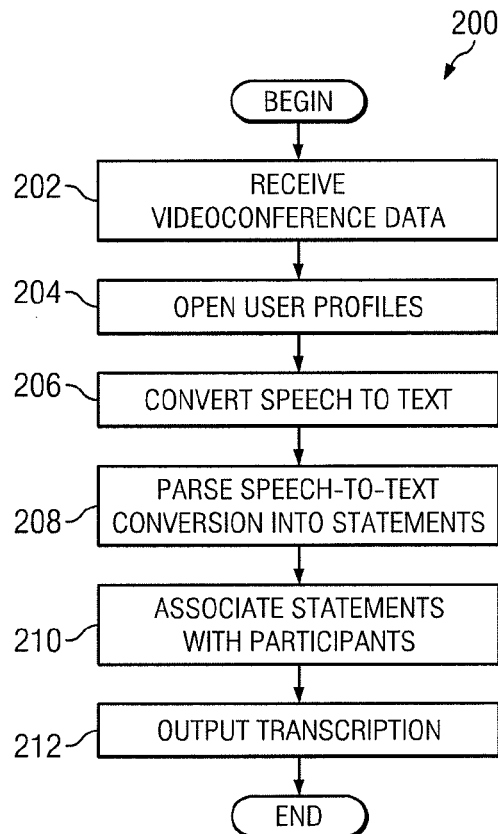
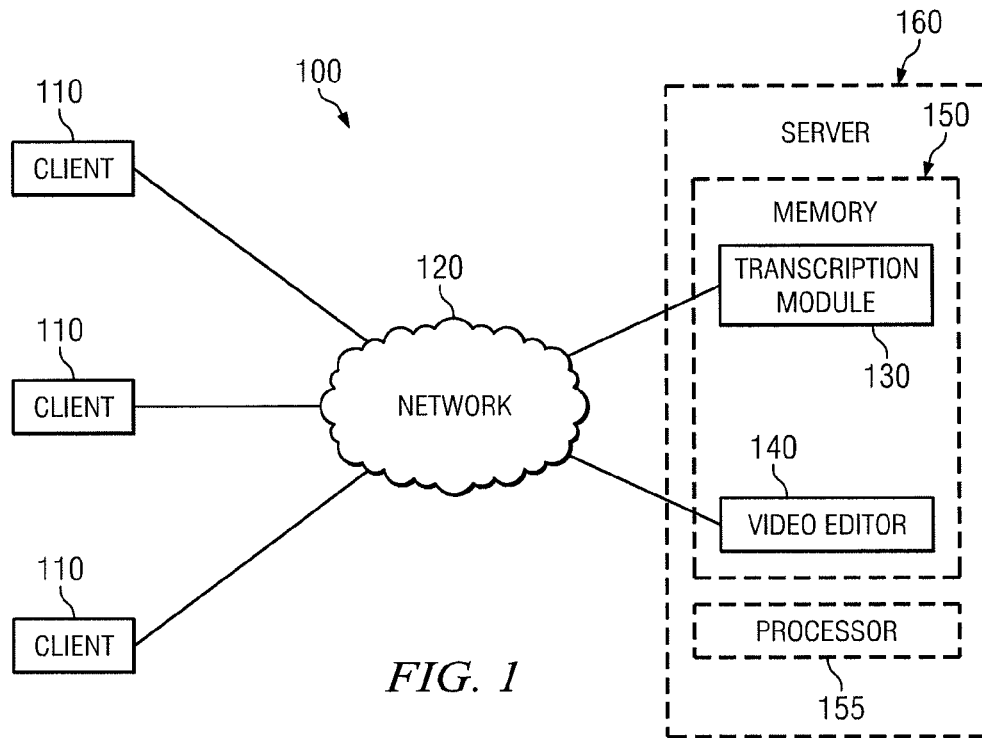
OTHER PUBLICATIONS

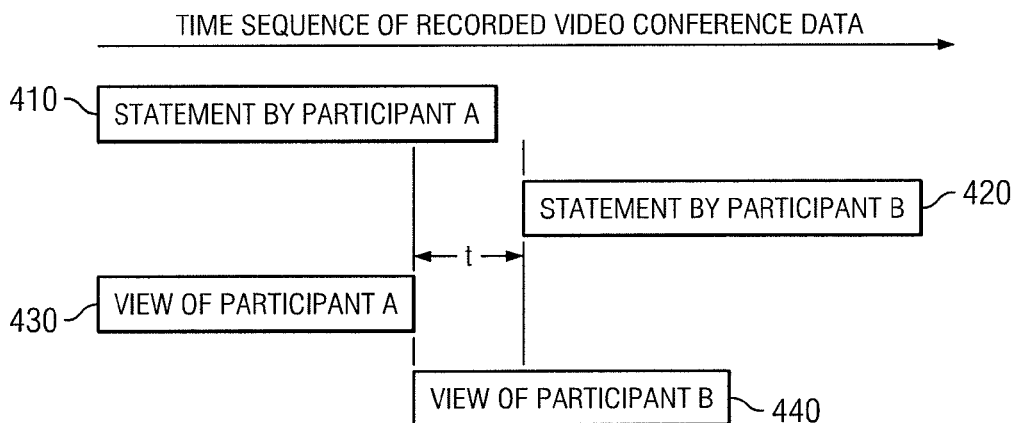
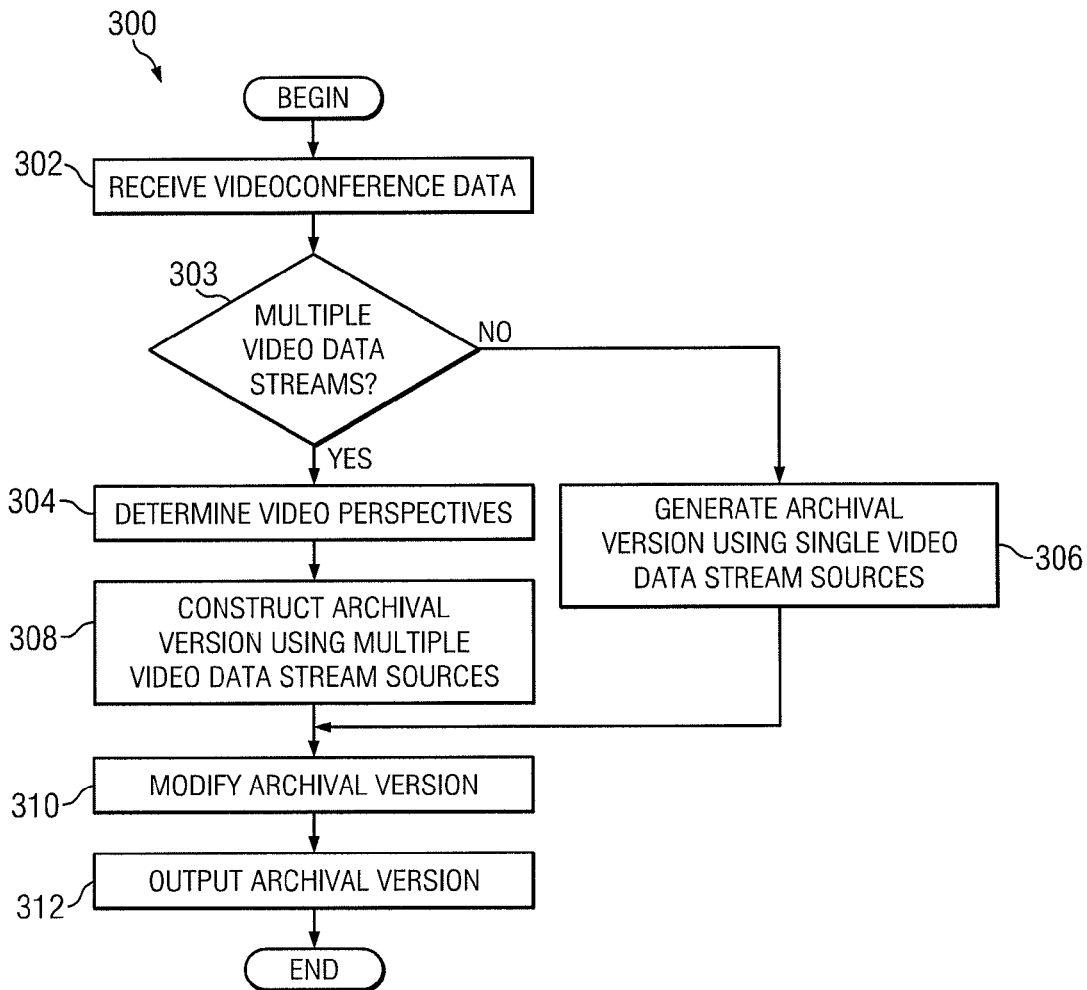
U.S. PATENT DOCUMENTS

| | | | | |
|--------------|------|---------|------------------------|-----------|
| 2004/0039464 | A1 * | 2/2004 | Virolainen et al. | 700/94 |
| 2006/0055771 | A1 * | 3/2006 | Kies | 348/14.03 |
| 2006/0251384 | A1 * | 11/2006 | Vronay et al. | 386/52 |
| 2009/0048939 | A1 * | 2/2009 | Williams | 705/26 |
| 2011/0217021 | A1 * | 9/2011 | Dubin et al. | 386/278 |

Risto Kurki-Suonio et al., A Method, A System and A Device for Converting Speech, U.S. Appl. No. 12/298,697, Filed May 28, 2009.
Michael A. Barasch et al., "Method and System for Providing Web Based Interactive Lessons with Improved Session Playback," U.S. Appl. No. 12/569,664, 159 pages, Filed Sep. 29, 2009.

* cited by examiner





1

SYSTEM AND METHOD FOR EDITING RECORDED VIDEOCONFERENCE DATA

TECHNICAL FIELD

This invention relates generally to the field of communications and more specifically to a system and method for editing recorded videoconference data.

BACKGROUND

Various videoconference (also known as a video teleconference) technologies exist that enable participants to interact simultaneously via video and audio transmissions. A videoconference may consist of a conversation between two people in private offices (point-to-point) or may involve multiple participants at various sites (multi-point). In addition to audio and visual transmission of various meeting activities, videoconferencing can be used to share documents, computer-displayed information, and whiteboards.

SUMMARY OF THE DISCLOSURE

In a method embodiment, a method for automatically editing data recorded during a videoconference includes accessing a plurality of video data streams. Each video data stream records a view of at least one of a plurality of human participants of the videoconference. The view recorded by each video data stream is different from the view recorded by each other video data stream. The method further includes determining, using one or more processors executing logic, that one of the plurality of video data streams recorded a view of a first one of the plurality of participants while the first one of the plurality of participants made a first statement. In addition the method includes determining, using one or more processors executing logic, that one of the plurality of video data streams recorded a view of a second one of the plurality of participants while the second one of the plurality of participants made a second statement after the first one of the plurality of participants made the first statement. An edited video data stream is generated using the plurality of video data streams. The edited video data stream comprises a transition that switches from the view of the first one of the plurality of participants to the view of the second one of the plurality of participants. The transition is timed such that when the edited video data stream is played the transition occurs before the commencement of the second statement.

Certain embodiments of the invention may provide one or more technical advantages. A technical advantage of one embodiment may be that audio-visual recordings of a videoconference may be automatically or semi-autonomously edited using a variety of criteria. In particular embodiments, multiple audio and visual data streams may be automatically edited into a single audio-visual data stream. Particular embodiments may generate archival versions of a videoconference that automatically switch the view from one speaker to another. Particular embodiments may also excise portions of the video and audio streams in order to abridge the videoconference recording—either to eliminate clearly irrelevant material, or to shorten the finished representation of the event in response to some user or machine specified request based upon content or time. Certain embodiments of the invention may include none, some, or all of the above technical advantages. One or more other technical advantages may be readily apparent to one skilled in the art from the figures, descriptions, and claims included herein.

2

BRIEF DESCRIPTION OF THE DRAWINGS

For a more complete understanding of the present invention and its features and advantages, reference is now made to the following description, taken in conjunction with the accompanying drawings, in which:

FIG. 1 illustrates a portion of a communication system according to one embodiment;

FIG. 2 is a flowchart illustrating a method for generating a transcription of a videoconference according to one embodiment;

FIG. 3 is a flowchart illustrating a method for generating an archival version of audio and visual data of a videoconference; and

FIG. 4 illustrates the use of a time sequence of data recorded during a videoconference.

DETAILED DESCRIPTION OF THE DRAWINGS

Embodiments of the present invention and its advantages are best understood by referring to FIGS. 1 and 2 of the drawings, like numerals being used for like and corresponding parts of the various drawings.

FIG. 1 illustrates a portion of a communication system 100 according to one embodiment. Communication system 100 generally includes multiple clients 110 communicatively coupled to a network 120. In certain embodiments, clients 110 and network 120 may cooperate together to enable one or more users to participate in a videoconference. Particular embodiments may include logic that facilitates recording information captured during videoconferences. For example, a transcription module 130 may be configured to generate speech-to-text transcriptions that identify various statements made during a videoconference in terms of both what was said and who most likely said it. As another example, a video editor 140 may be configured to generate an archival audio-video data stream of a videoconference that switches between the differing viewing perspectives of multiple video data streams recorded during the videoconference.

Clients 110 may include devices that end users or other devices may use to initiate or participate in a videoconference. For example, clients 110 may include a computer, a personal digital assistant (PDA), a laptop, an electronic notebook, a telephone, a mobile station, an audio IP phone, a video phone appliance, a personal computer (PC) based video phone, a streaming client, or any other device, component, element, or object capable of engaging in voice, video, and/or data exchanges within videoconference system 100.

Clients 110 may include a suitable interface to a human user. For example, clients 110 may include a microphone, a video camera, a display, a keyboard, a whiteboard, any combination of the preceding, or other terminal equipment that may provide a videoconferencing interface. Various client 110 interfaces may be configured to capture various forms of data of a videoconference and communicate the captured data to network 120 in the form of a data stream. Data, as used herein in this document, refers to any type of numeric, voice and audio, video, audio-visual, or script data, or any type of source or object code, any combination of the preceding, or any other suitable information in any appropriate format that may be communicated from one point to another.

In particular embodiments, client 110 interfaces may enable a user who did not actively participate in a videoconference to review an edited audio-visual recording of the videoconference. For example, client 110 interfaces may enable a non-participating user to watch an edited version of a videoconference while the videoconference is in progress

and system 100 edits data in real-time. Under this scenario, system 100 may broadcast to one or more clients 110 a live or near-live recording of the videoconference edited by system 100. Alternatively, client 110 interfaces may enable a non-participating user to watch an edited version of a videoconference after the videoconference has terminated and system 100 has effected all data processing.

Network 120 may comprise any wireless network, wireline network, or combination of wireless and wireline networks capable of supporting communication of data. For example, network 120 may include all or a portion of a public switched telephone network (PSTN), a public or private data network, a local area network (LAN), a metropolitan area network (MAN), a wide area network (WAN), a local, regional, or global communication or computer network such as the Internet, a wireline or wireless network, an enterprise intranet, other suitable communication link, or any combination of the preceding. In a particular embodiment, network 120 may include a centralized system capable of supporting videoconferencing by receiving media streams from particular clients 110 connected to the same videoconference session, mixing the streams, and sending individual streams back to those clients 110.

Transcription module 130 may include any suitable logic configured to generate speech-to-text transcriptions of videoconferences. Certain speech-to-text transcriptions generated by transcription module 130 may identify one or more respective participants as the likely source of various statements made during the videoconference, as explained further below with reference to FIG. 2. In certain embodiments, the operations of transcription module 130 may be performed using any suitable logic comprising software, hardware, and/or other logic.

Video editor 140 may be configured to generate an archival audio-video data stream of a videoconference. In certain embodiments, video editor 140 may use a variety of rules to switch between the differing viewing perspectives of multiple video data streams recorded during the videoconference, as explained further below with reference to FIG. 3. In certain embodiments, at least a portion of the operations of video editor 140 may be performed in real time as a videoconference progresses. In various embodiments, at least a portion of the operations of video editor 140 may be performed after the videoconference has concluded. In particular embodiments, the operations of video editor 130 may be performed using any suitable logic comprising software, hardware, and/or other logic.

In certain embodiments, transcription module 130 and/or video editor 140 may include logic stored in computer-readable memory 150. Memory 150 stores information. A memory 150 may comprise one or more tangible, computer-readable, and/or computer-executable storage media. Examples of memory 150 include computer memory (for example, Random Access Memory (RAM) or Read Only Memory (ROM)), mass storage media (for example, a hard disk), removable storage media (for example, a Compact Disk (CD) or a Digital Video Disk (DVD)), database and/or network storage (for example, a server), and/or other computer-readable medium. Although FIG. 1 illustrates transcription module 130 and video editor 140 as residing at the same memory 150, in alternative embodiments transcription module 130 and video editor 140 may reside at separate memory 150 with respect to each other. In particular embodiments, transcription module 130 and/or video editor 140 may reside at one or more memory devices 150 accessible to or through one or more servers 160.

Server 160 generally refers to any suitable device capable of communicating with client 110 through network 120. For example, server 160 may be a file server, a domain name server, a proxy server, a web server, an application server, a computer workstation, a handheld device, one or more other devices operable to communicate with client 102, or any combination of the preceding. In some embodiments, server 160 may restrict access only to a private network (e.g. a corporate intranet); or, in some other embodiments, server 160 may publish pages on the World Wide Web. In this example, server 160 generally includes at least memory 150 and one or more processors 155; however, any suitable server(s) 160 may be used. Although FIG. 1 illustrates memory 150 residing within server 160, all or a portion of memory 150 may alternatively reside separate from server 160.

Particular embodiments may be implemented as a software as a service ("SaaS"). For example, a provider may license transcription module 130 and/or video editor 140 to customers as a service on demand, through a subscription model, a "pay-as-you-go" model, and/or through some other payment model. As another example, customers may be granted access to, and/or control of, transcription module 130 and/or to video editor 140 via network 120 for purposes of generating transcriptions of their videoconferences.

FIG. 2 is a flowchart 200 illustrating a method for generating a transcription of a videoconference according to one embodiment. In step 202, data regarding the videoconference is received. For example, transcription module 130 may receive audio data that includes an acoustic encoding of human speech and/or other auditory input captured from the videoconference. As another example, transcription module 130 may receive video data that includes an encoding of images and/or other visual sensory information captured from the videoconference.

In a particular embodiment, data may be received in step 202 in the form of one or more data streams. A data stream may be comprised of a variety of different data types and/or data combinations from various sources. For example, each client 110 facilitating the same videoconference may generate respective audio and visual data streams, thereby providing multiple client-based perspectives. As another example, a data stream may be comprised of a combination of data provided by two or more clients 110 facilitating the videoconference. In certain instances, audio and video data captured by a particular client 110 may be received as separate audio and video data streams, respectively. Alternatively, certain audio and video data captured by a particular client 110 may be received as a combined audio-visual data stream.

In various embodiments, data may be received in real time in step 202 as the data is captured by one or more clients 110 during the videoconference. In alternative embodiments, some or all of the data may be received in step 202 sometime after the videoconference has terminated. For example, data may be uploaded or downloaded in step 202 from computer-readable memory.

In step 204, a user profile is opened for each human participant of the videoconference. In certain embodiments, a user profile may comprise data that uniquely identifies the user. For example, the user profile data may identify the user's voice profile, speech recognition profile, the user's facial features, the user's location in a room or building, the site at which the user is participating in the videoconference, an address (e.g., electronic and/or physical) of a client 110 in use by the user, any combination of the preceding, or other information that may be used to establish a profile that identifies the user from among those participating in the video conference.

5

ence. In certain instances, a user profile may comprise data that is determined prior to a videoconference in which the user is participating, during the videoconference, and/or after the videoconference concludes.

In certain embodiments, the step of opening a user profile may include retrieving, creating and/or modifying the user profile. For example, data captured during the videoconference in step 202, or during a prior videoconference, may be used to create or modify user profile data identifying the user's voice profile and/or the user's facial features. As another example, a user may be asked to state a series of words. The sound of the user's voice in stating those words may then be used to define or redefine a voice profile for the user, which may be recorded as user profile data.

In step 206, human speech of the videoconference is converted into symbolic form. For example, transcription module 130 may use data captured from the videoconference and a set of extract rules to convert human speech into text.

In step 208, at least portions of the converted human speech may be parsed into individual statements. For example, transcription module 130 may make determination that a collection of spoken words or other sounds likely came from a particular sound source or from a collection of sound sources. Based at least in part on this determination, transcription module 130 may logically identify this collection of audio data as a statement.

In step 210, each statement is associated with one or more sound sources of the videoconference. In certain embodiments, the association may be at least partially effected automatically by transcription module 130. For example, transcription module 130 may determine which participant likely spoke the statement by matching human speech of a statement to a voice profile of a particular participant, analyzing video data to determine which participant's facial movement appears to be synchronized with audio data of the statement, determining the source of a data stream corresponding to the statement, any combination of the proceeding, or by any of a variety of other methods including textual and semantic analysis.

In various embodiments, transcription module 130 may determine confidence levels representing the probabilities that one or more participants are the sources of a particular statement. Transcription module 130 may set an alert and/or perform additional analysis if no participant is attributed a probability greater than a predetermined threshold. For example, using voice profiles alone transcription module 130 may determine the probability that either participant A or participant B made a particular statement is 70% and 30%, respectively. If the maximum confidence level determined for a particular statement does not exceed the predetermined threshold, transcription module may perform additional analysis, such as analysis involving facial movement, in an attempt to increase the maximum confidence level.

In certain instances, a lower maximum confidence level may trigger transcription module 130 to enable human-assisted transcription. In this mode, a sound clip, a video clip, transcribed text, and/or other data corresponding to the statement may be presented to a human reviewer. The human reviewer may then be prompted to select the source of the statement from among all the participants of the videoconference or from a subset of participants selected by the transcription module.

In step 212, a transcription output is generated that identifies statements of the videoconference and respective sources for those statements. The transcription output may be in any suitable form including, for example, in printed form and/or in computer-readable form. Certain computer-readable forms

6

may be suitable for downloading, printing, performing a text-based search, for wireless or wireline transmission, and/or for storage in computer-readable media.

FIG. 3 is a flowchart 300 illustrating a method for generating an archival version of audio and visual data of a videoconference. For particular videoconferences, audio and/or visual data may be recorded from multiple, differing perspectives that are synchronized together. For example, each client 110 used to facilitate a videoconference may be configured to record one or more respective audio and/or video data streams during the videoconference. In certain embodiments, video editor 140 may use a variety of computer-implemented rules to generate a master archival audio-video data stream that switches between the different available audio and/or visual perspectives recorded during the videoconference. The master archival audio-video data stream generated by video editor 140 may be sufficiently representative of the videoconference, such that it may not be necessary to also archive all of the available audio and/or visual perspectives used to generate the archival version. The master archival audio-video data stream may be sufficiently representative so that people later trying to understand what happened at a video conference may not need to refer back to the original 'raw footage'.

In step 302, data regarding the videoconference is received. In various embodiments, the data may be received in a manner substantially similar to certain examples described previously with reference to step 202 of FIG. 2. If video editor 140 determines in step 303 the data received in step 302 includes multiple, synchronized video data streams of the videoconference recorded from different visual perspectives, flowchart 300 proceeds to step 304; otherwise, an archival version of the videoconference is generated in step 306 using the single video data stream recorded for the videoconference.

In step 304, video editor 140 may determine which perspective or combination of perspectives of multiple, synchronized video data streams to include in each sequence of temporally-ordered video frames of the archival version of the videoconference. Any of a variety of criteria may be used in making the determination. For example, the determination for any given video frame may be based on which videoconference participant is speaking, which participant is the next to speak, which participant is considered the most important speaker during a video frame when multiple participants are speaking at once, any combination of the preceding, or other suitable criteria.

If the criterion in step 304 is based in part on who is speaking during a video frame, video editor 140 may determine who is speaking during the video frame using data generated by transcription module 130. In an alternative embodiment, video editor 140 may make a determination as to who spoke during a particular video frame in a manner substantially similar to that described previously with reference to step 210 above.

In particular instances, video editor 140 may select in step 304 a combination of perspectives for a particular video frame sequence. For example, video editor 140 may edit two different video data streams into a combined, split-screen video frame sequence in response to a determination that multiple participants are speaking at once, in response to a determination that multiple participants are speaking in rapid succession, and/or in response to a determination that the viewers' interests would be best served by utilizing this format for any reason (including making non-verbal responses visible).

In certain embodiments, video editor 140 may edit in step 304 a particular video sequence of the archival version of a videoconference in a manner that shows a view of one or more

7

participants at moments during the videoconference when another participant was speaking. As shown in FIG. 4, for example, system 100 may determine from recorded videoconference data that participant A made a statement during a first time sequence 410 and that participant B made the next statement during a subsequent time sequence 420 of the recorded videoconference. Based at least in part on this determination, video editor 140 may transition the view of the archival version of the videoconference from a view 430 of participant A to a view 440 of participant B before time sequence 410 terminates (i.e. while participant A was speaking) and before time sequence 420 begins (i.e. before participant B spoke the next statement). As shown in FIG. 4, t represents the time interval during which the archival version of the videoconference will show a view of participant B while participant B is not speaking.

As another example of intelligent processing, system 100 may choose which participant to show during a recorded timeframe when participant A and participant B spoke simultaneously. The decision may be based, for example, on a determination of which speaker is more important and/or a determination of which speaker is speaking more on topic. In certain instances, there may be a time gap between a statement made by participant A and a subsequent statement made by participant B. System 100 may be configured to show both participant A and participant B during the gap time period, or the gap time period can be split with some time showing participant A and some time showing participant B.

As yet another example of intelligent processing performed in step 304 that may result in not showing a view of a participant while the participant spoke, a time sequence of recorded videoconference data may correspond to a longer statement or a series of statements spoken by a particular participant that is intermittently interrupted by short statements, verbal acknowledgements, or other sounds (e.g., laughter, coughing, shuffling, etc.) made by other participants. Video editor 140 may determine those types of short or intermittent interruptions are not significant enough to switch the perspective away from the more important participant speaking the longer statement or series of statements. This type of intelligent decision making may be contrasted with alternative systems that switch the perspective of a video stream based on sound sources alone, which may result in choppy and visually irritating video cuts.

In still another example of intelligent processing performed in step 304 that may result in not showing a view of a participant while the participant spoke during the videoconference, video editor 140 may analyze semantics of the videoconference transcription to intelligently determine which view or combination of views of the videoconference to use during a particular time sequence when multiple participants are speaking at once. For example, video editor 140 may search the multiple statements for key words spoken with particular frequency during the videoconference to determine who is speaking on topic and who is having an aside about something unrelated to the subject matter of the videoconference. Thus, system 100 may look at the meaning of statements made during the videoconference to intelligently select which view or combination of views to use. Statistical analysis, human-specified agendas, and/or other input may be used to assist in identifying various key words that may be considered on topic for a particular videoconference. This type of intelligent decision making may be contrasted with alternative systems that switch the perspective of a video stream based merely on who is talking at any given point in time.

Semantic analysis may also be performed in step 304 to determine which time sequences of the recorded videocon-

8

ference data to include and which to discard in an archival version of the videoconference that is limited to highlights. Such processing may be effected by removing portions of the recorded videoconference data that are semantically unrelated to key words as determined in a manner substantially similar to that discussed above. In certain embodiments, system 100 may receive input from a viewer-user and create a customized archival version of the videoconference that is limited to particular highlights associated with the input specified by the viewer-user.

In step 308, video editor 140 may construct an archival version of the videoconference based at least in part on the determinations made in step 304. In step 310, video editor 140 may modify the archival version constructed in step 308 based on input received from one or more users. In step 312, the constructed archival version of the videoconference may be outputted in a manner that may be suitable for downloading, printing, performing a text-based search, for wireless or wireline transmission, and/or for storage in computer-readable media.

The archival version may include many types of metadata associated with various views and time points of the videoconference. These may include who is speaking when, keywords associated with different periods of the conference, gestural and emotion analysis of participants, and so on. Portions of this metadata may be derived using computations based upon the explicit spoken content of the videoconference. Other portions may be derived from temporal dynamics of the interactions and nonverbal communications that the system may be able to note and/or interpret.

In certain embodiments, video editor 140 may use a variety of computer-implemented rules to construct the audio portion of an archival version of a videoconference in addition to the video portion. For example, video editor 140 switch between differing recorded audio perspectives based on which audio data stream has the highest fidelity or quality at any given point in time. Any of a variety of factors may influence the quality of portions of an audio data stream. For example, a microphone directly recording human speech of a participant may produce better audio quality than that produced by a microphone recording the same human speech as produced by a speaker. As another example, video editor 140 may switch to the audio data stream with the least noise anomalies, independent of who is speaking.

Modifications, additions, or omissions may be made to the systems and apparatuses disclosed herein without departing from the scope of the invention. The components of the systems and apparatuses may be integrated or separated. For example, network 120 may include transcription module 130 and/or video editor 140. Moreover, the operations of the systems and apparatuses may be performed by more, fewer, or other components. For example, the operations of a particular client 110 and transcription module 130 may be performed by one component, or the operations of transcription module 130 and/or video editor 140 may be performed by more than one component. In addition, one or more forms of logic may be configured to perform the operations of both transcription module 130 and video editor 140. Operations of the systems and apparatuses may be performed using any suitable logic comprising software, hardware, and/or other logic.

Modifications, additions, or omissions may be made to the methods disclosed herein without departing from the scope of the invention. The methods may include more, fewer, or other steps. Additionally, steps may be performed in any suitable order, steps sequences may loop, and certain steps may be

repeated. For example, a user profile may be opened in step 204 before data is received in step 202.

A component of the systems and apparatuses disclosed herein may include an interface, logic, memory, and/or other suitable element. An interface receives input, sends output, processes the input and/or output, and/or performs other suitable operation. An interface may comprise hardware and/or software. Logic performs the operations of the component, for example, executes instructions to generate output from input. Logic may include hardware, software, and/or other logic. Logic may be encoded in one or more tangible media and may perform operations when executed by a computer. Certain logic, such as a processor, may manage the operation of a component. Examples of a processor include one or more computers, one or more microprocessors, one or more applications, and/or other logic.

In particular embodiments, the operations of the embodiments may be performed by one or more computer readable media encoded with a computer program, software, computer executable instructions, and/or instructions capable of being executed by a computer. In particular embodiments, the operations of the embodiments may be performed by one or more computer readable media storing, embodied with, and/or encoded with a computer program and/or having a stored and/or an encoded computer program.

Although this disclosure has been described in terms of certain embodiments, alterations and permutations of the embodiments will be apparent to those skilled in the art. Accordingly, the above description of the embodiments does not constrain this disclosure. Other changes, substitutions, and alterations are possible without departing from the spirit and scope of this disclosure, as defined by the following claims.

What is claimed is:

1. A method for automatically editing data recorded during a videoconference, comprising:

accessing a plurality of video data streams, each video data stream recording a view of at least one of a plurality of human participants of the videoconference, the view recorded by each video data stream being different from the view recorded by each other video data stream;

determining, using one or more processors executing logic, that one of the plurality of video data streams recorded a view of a first one of the plurality of participants while the first one of the plurality of participants made a first statement;

determining, using one or more processors executing logic, that one of the plurality of video data streams recorded a view of a second one of the plurality of participants while the second one of the plurality of participants made a second statement after the first one of the plurality of participants made the first statement; and

generating an edited video data stream using the plurality of video data streams, the edited video data stream comprising a transition that switches from the view of the first one of the plurality of participants to the view of the second one of the plurality of participants, the transition timed such that when the edited video data stream is played the transition occurs before the commencement of the second statement, wherein the transition is based at least in part on a determination that one of the plurality of participants is more important than another one of the plurality of participants, and wherein the determination that one of the plurality of participants is more important than another one of the plurality of participants is based at least in part on which participant of the plurality of

participants contribute the most human speech content during the entire videoconference.

2. The method of claim 1, further comprising determining when, relative to one of the plurality of video data streams, the first statement ends.

3. The method of claim 2, wherein the determination of when the first statement ends is based at least in part on one or more audio data streams recorded during the videoconference.

4. The method of claim 2, wherein the determination of when the first statement ends is based at least in part on a determination that audio data of one or more audio data streams recorded during the videoconference comprises an audio interruption made by one of the plurality of participants during the first statement, the end of the first statement occurring after the audio interruption.

5. The method of claim 1, further comprising determining when, relative to one of the plurality of video data streams, the second statement begins.

6. A non-transitory computer-readable memory storing logic, the logic operable when executed by one or more processors to:

access a plurality of video data streams, each video data stream recording a view of at least one of a plurality of human participants of the videoconference, the view recorded by each video data stream being different from the view recorded by each other video data stream;

determine that one of the plurality of video data streams recorded a view of a first one of the plurality of participants while the first one of the plurality of participants made a first statement;

determine that one of the plurality of video data streams recorded a view of a second one of the plurality of participants while the second one of the plurality of participants made a second statement after the first one of the plurality of participants made the first statement; and

generate an edited video data stream using the plurality of video data streams, the edited video data stream comprising a transition that switches from the view of the first one of the plurality of participants to the view of the second one of the plurality of participants, the transition timed such that when the edited video data stream is played the transition occurs before the commencement of the second statement, wherein the transition is based at least in part on a determination that one of the plurality of participants is more important than another one of the plurality of participants, and wherein the determination that one of the plurality of participants is more important than another one of the plurality of participants is based at least in part on which participant of the plurality of participants contribute the most human speech content during the entire videoconference.

7. The non-transitory computer-readable memory of claim 6, wherein the logic is further operable when executed by the one or more processors to determine when, relative to one of the plurality of video data streams, the first statement ends.

8. The non-transitory computer-readable memory of claim 7, wherein the determination of when the first statement ends is based at least in part on one or more audio data streams recorded during the videoconference.

9. The non-transitory computer-readable memory of claim 7, wherein the determination of when the first statement ends is based at least in part on a determination that audio data of one or more audio data streams recorded during the videoconference comprises an audio interruption made by one of

11

the plurality of participants during the first statement, the end of the first statement occurring after the audio interruption.

10. The non-transitory computer-readable memory of claim **6**, wherein the logic is further operable when executed by the one or more processors to determine, relative to one of the plurality of video data streams, the second statement begins.

11. A system comprising:

one or more processors; and

computer-readable memory comprising logic, the logic operable when executed by the one or more processors to:

access a plurality of video data streams, each video data stream recording a view of at least one of a plurality of human participants of the videoconference, the view recorded by each video data stream being different from the view recorded by each other video data stream;

determine that one of the plurality of video data streams recorded a view of a first one of the plurality of participants while the first one of the plurality of participants made a first statement;

determine that one of the plurality of video data streams recorded a view of a second one of the plurality of participants while the second one of the plurality of participants made a second statement after the first one of the plurality of participants made the first statement; and

generate an edited video data stream using the plurality of video data streams, the edited video data stream com-

12

prising a transition that switches from the view of the first one of the plurality of participants to the view of the second one of the plurality of participants, the transition timed such that when the edited video data stream is played the transition occurs before the commencement of the second statement, wherein the transition is based at least in part on a determination that one of the plurality of participants is more important than another one of the plurality of participants, and wherein the determination that one of the plurality of participants is more important than another one of the plurality of participants is based at least in part on which participant of the plurality of participants contribute the most human speech content during the entire videoconference.

12. The system of claim **11**, wherein the logic is further operable when executed by the one or more processors to determine when, relative to one of the plurality of video data streams, the first statement ends.

13. The system of claim **12**, wherein the determination of when the first statement ends is based at least in part on one or more audio data streams recorded during the videoconference.

14. The system of claim **12**, wherein the determination of when the first statement ends is based at least in part on a determination that audio data of one or more audio data streams recorded during the videoconference comprises an audio interruption made by one of the plurality of participants during the first statement, the end of the first statement occurring after the audio interruption.

* * * * *